非正則分布族のフィンスラー幾何学(Non-Regular Statistical Estimation)

AUTHOR(S):

甘利, 俊一

数理解析研究所講究録
538巻 1984年 81-95

非正則分布族のフィンスラー幾何学

東大工学部　甘利俊一 (Shun-ichi Amari)

1. はじめに

本報告は，非正則な統計分布の族が幾何学的にはフィンスラー空間の構造を持つことを指摘し，非正則なモデルのもとでの統計的推論の漸近理論がフィンスラー幾何学を用いて建設できることを主張せんとするものである。これにはベクトル確率変数の安定分布が Minkowski 計量を持つことが関係している。本稿はこの問題に関する予備的考察であって，理論として定式化できているものではないことを断っておく。

2. Minkowski 計量と Finsler 空間

n 次元の線形空間 T を考えよう。T の要素であるベクトル x は基底 $\{e_i\}$ $(i=1, \cdots, n)$ を用いて

$$x = x^i e_i$$

のように表わせる（ここで指標 i についての和をとる Einsteinの簡約和法を用いる)。このとき，xを成分x^i ($i=1,\cdots,n$)で表わすことができる。線形空間Tに，ベクトルxの長さを表わす基本関数

$$\|x\| = F(x)$$

が与えられているものとしよう。Fは1次の斉次関数，すなわち

$$F(cx) = cF(x), \qquad c>0$$

を満たす凸関数であるとし，

$$F(x) \geqq 0$$

で等号は$x=0$のときこのときにのみ成立するものとする。さらに，ここでは$F(-x)=F(x)$も仮定する。このようなノルムを定義する基本関数Fが与えられている線形空間をMinkowski空間と呼ぶ。ユークリッド空間はF^2が二次形式

$$F^2(x) = g_{ij}x^i x^j,$$

g_{ij}は正定の定数テンソル，で与えられるMinkowski空間である。

Tにおいて

$$F(x) = 1$$

を満たすxの集合（これは原点を囲む$n-1$次元曲面となる）をindicatrixと呼ぶ。$F(x) \leqq 1$で定義される領域は

凸である。ユークリッド空間とは，indicatrix が二次曲面（楕円体）となる空間のことである。

多様体Sを考え，Sの局所座標系を $\theta = (\theta^1, \cdots, \theta^n)$ としよう。一点 θ におけるSの接空間を T_θ とおく。各 T_θ に基本関数 $F(x, \theta)$ があたえられている多様体，すなわち接空間が Minkowski 空間であるような多様体を Finsler 空間という。（Fの微分可能性を仮定する。）F^2 が x の二次関数となる Riemann 空間は Finsler 空間の特殊な場合である。Riemann 空間の場合と同様に，Finsler 空間の曲線にそった Fの積分によって，曲線 $\theta(s)$ の長さを

$$\int F(\theta, \dot{\theta}(s))ds$$

で定義できる（$\dot{\theta} = d\theta/ds$）。また，ここでは述べないが接空間の接続を導入することができる。

Minkowski 空間Tに話を戻そう。Fは1次の斉次関数だから，$F^2(x) = \{F(x)\}^2$ として

$$g_{ij}(x) = \frac{1}{2}\frac{\partial^2}{\partial x^i \partial x^j} F^2(x)$$

は0次の斉次関数を成分とする正定行列となる。すなわち

$$g_{ij}(cx) = g_{ij}(x)$$

さらに

$$F^2(x) = g_{ij}(x)\, x^i x^j$$

が成立する。

例　L_p 空間

$$F(x) = \{|x^1|^p + |x^2|^p\}^{\frac{1}{p}}$$

のときは

$$g_{11}(x) = (2-p)|x^1|^{2p-1}\{|x^1|^p + |x^2|^p\}^{\frac{2}{p}-2} + (p-1)|x^1|^{p-2}\{|x^1|^p + |x^2|^p\}^{\frac{2}{p}-1}$$

$$g_{12}(x) = |x^1|^{p-1}|x^2|^{p-1}\left(\frac{2}{p}-1\right)p\,\{|x^1|^p + |x^2|^p\}^{\frac{2}{p}-2}$$

などとなる。

二つのベクトル x, y に対して

$$g_{ij}(x)\, x^i y^j = 0$$

が成立するとき, y は x に対して直交するという。このとき x は y に直交するとは限らない。

ベクトル x に対して双対ベクトル x^* を次式で定義しよう。

$x = (x^i)$, $x^* = (x_i)$ として

$$x_i = g_{ij}(x)\, x^j.$$

このとき

$$F^2(x) = x^i x_i$$

が成立する。

x と x^* は Legendre 変換で結ばれている。すなわち，

$$\mathfrak{u} = x^* \qquad (u_i = x_i)$$

として，関数 $H(\mathfrak{u})$ が存在して

$$u_i = \frac{1}{2}\frac{\partial F^2(x)}{\partial x^i} , \qquad x^i = \frac{1}{2}\frac{\partial H^2(\mathfrak{u})}{\partial u_i}$$

の関係が成立し，

$$F^2(x) + H^2(\mathfrak{u}) - 2x^i u_i = 0$$

の関係で結ばれている。$H(\mathfrak{u})$ は1次の斉次関数

$$H(c\mathfrak{u}) = |c|\, H(\mathfrak{u})$$

で，

$$F(x) = H(\mathfrak{u})$$

が成立する。$H(\mathfrak{u})$ は双対空間 $T^* = \{\mathfrak{u}\}$ に Minkowski 計量を定義する。

$$g^{ij}(\mathfrak{u}) = \frac{1}{2}\,\frac{\partial^2 H^2(\mathfrak{u})}{\partial u_i \partial u_j}$$

とすると，

$$H^2(\mathfrak{u}) = g^{ij}(\mathfrak{u})\, u_i u_j$$

であって，$g_{ij}(x)$ と $g^{ij}(\mathfrak{u})$ とは互に逆行列の関係で結ばれ

$$x^i = g^{ij}(\mathfrak{u})\, u_j , \qquad u_i = g_{ij}(x)\, x^j$$

である。

3. 安定分布と Minkowski 計量

いま T^* に値をとる確率変数 $\boldsymbol{z} = (z_i)$　$(i=1,2,\cdots,n)$ を考よう。同一で独立な分布に従がう任意の k 個の確率変数 $\boldsymbol{z}_1, \boldsymbol{z}_2, \cdots, \boldsymbol{z}_k$ に対してある a_k が存在して

$$\frac{S_k}{a_k} = \frac{1}{a_k}(\boldsymbol{z}_1 + \cdots + \boldsymbol{z}_k)$$

が元の $\boldsymbol{x}$ と同一の分布に従がうとき，この分布を安定分布という。分布の特性関数 $\varphi(\boldsymbol{t})$，$\boldsymbol{t} = (t^i)$，$i=1,\cdots,n$ を

$$\varphi(\boldsymbol{t}) = E[\exp\{i\,\boldsymbol{t}\cdot\boldsymbol{z}\}], \quad \boldsymbol{t}\cdot\boldsymbol{z} = t^i z_i$$

で定義すると，安定分布は特性関数を用いて

$$[\varphi(\boldsymbol{t})]^k = \varphi(a_k \boldsymbol{t})$$

で定義できる。$a_k = k^{1/\alpha}$，　$1 \leqq \alpha \leqq 2$ であることが知られていて，$a_k = k^{1/\alpha}$ となる分布を指数 α の安定分布という。キュムラント関数 $\psi(\boldsymbol{t})$

$$\psi(\boldsymbol{t}) = \log \varphi(\boldsymbol{t}), \quad \varphi(\boldsymbol{t}) = e^{\psi(\boldsymbol{t})}$$

を用いると，$\psi(\boldsymbol{t})$ は

$$\psi(\boldsymbol{t}) = -A(\boldsymbol{t})\{1 + i\,G(\boldsymbol{t})\}$$

と書ける（Lévy-Khinchin 表示の一般化）。ここに，$\psi(\boldsymbol{t})$

は α次の斉次正関数

$$A(c\mathbf{t}) = |c^{\alpha}|\, A(\mathbf{t}),$$

$G(\mathbf{t})$ は 0次の斉関数で

$$G(-\mathbf{t}) = -G(\mathbf{t})$$

である。ここでは簡単のため，$\mathbf{x}$ と $-\mathbf{x}$ の分布は等しいと仮定する。すると

$$\Psi(\mathbf{t}) = -A(\mathbf{t}), \qquad \varphi(\mathbf{t}) = \exp\{-A(\mathbf{t})\}$$

となる。

ここで $\mathbf{t}$ の空間 T に Minkowski 計量を

$$F(\mathbf{t}) = \{A(\mathbf{t})\}^{1/\alpha}$$

により導入する。すると計量テンソルは

$$g_{ij}(\mathbf{t}) = \frac{1}{2}\frac{\partial^2 F^2}{\partial t^i \partial t^j}$$

であたえられる。かくして，安定分布に対して Minkowski空間が一つ対応する。とくに $\alpha = 2$ の場合は安定分布は正規分布となり，$g_{ij}(\mathbf{t})$ は $\mathbf{t}$ によらないから二次形式 g_{ij} による計量が一つ定義できる。明らかに z_i は $N(0, g_{ij})$ に従がう正規分布となる。一般の場合には，

$$\varphi(\mathbf{t}) = \exp\left\{-\left[g_{ij}(\mathbf{t})\, t^i t^j\right]^{\frac{\alpha}{2}}\right\}$$

のように書ける。

議論を双対に進めると，確率変数 $\mathbf{z}=(z_i)$ の空間に

$$\|\mathbf{z}\| = H(\mathbf{z})$$

によって Minkowski ノルムを導入できる。 H は F より Legendre 変換によって得られる。

$$g^{ij}(\mathbf{z}) = \frac{1}{2}\frac{\partial^2 H(\mathbf{z})}{\partial z_i \partial z_j}$$

で，

$$\|\mathbf{z}\|^2 = \frac{1}{2} g^{ij}(\mathbf{z}) z_i z_j$$

である。 $\mathbf{z}$ と $\mathbf{t}$ とを

$$t^i = g^{ij}(\mathbf{z}) z_j \quad , \qquad z_i = g_{ij}(\mathbf{t}) t^j$$

で結びつけると，$\mathbf{t}$ を確率変数と見たし空間 T に確率測度を導入できる。 $\mathbf{t}$ の分布を調べる必要がある。

いま，二つの指数 α の定常分布 $\mathbf{z}_1$, $\mathbf{z}_2$ があるとき，その直積（直和）を $\mathbf{z}=(\mathbf{z}_1, \mathbf{z}_2)$ とする。 $\mathbf{z}_1$ と $\mathbf{z}_2$ は確率変数として独立であるとする。このとき， $\mathbf{t}=(\mathbf{t}_1, \mathbf{t}_2)$ を対応する双対ベクトルとして，$\mathbf{z}$ の特性関数は

$$\varphi(\mathbf{t}) = E[e^{i\mathbf{z}\cdot\mathbf{t}}] = E[e^{i\mathbf{z}_1\cdot\mathbf{t}_1}]\,E[e^{i\mathbf{z}_2\cdot\mathbf{t}_2}]$$

となる。したがって $\mathbf{t}$ の基本関数 $F(\mathbf{t})$ は

$$F(\mathbf{t}) = \left[\{F_1(\mathbf{t}_1)\}^\alpha + \{F_2(\mathbf{t}_2)\}^\alpha\right]^{\frac{1}{\alpha}},$$

ここに F_1, F_2 は $\mathcal{H}_1$, $\mathcal{H}_2$ の基本関数である。これは指数 α の安定分布の L_α 空間としての性格を明らかにする。このとき双対である $H(z)$ は $L_{\alpha'}$ 空間 $\{(1/\alpha)+(1/\alpha')=1\}$ の性質を持つと考えられる。

4. 非正則統計的モデルと Finsler 空間

n 次元空間の translation family として，分布密度関数

$$p(x, \theta) = f(x-\theta)$$

確率変数の分布族 $S=\{p(x,\theta)\}$ を考える。ここに $\theta=(\theta^1, \cdots, \theta_n)$ は分布を指定するパラメータであり，関数 $f(x)$ の台は compact な凸領域で，その境界における立ち上がりは，境界からの距離 d に応じて

$$f(x) = c\,d^{\alpha-1}, \qquad 1 \leqq \alpha \leqq 2$$

の形をしているものとする。1次元のときは，たとえば

$$f(x, \theta) = C_\alpha (x-\theta)^{\alpha-1}(1-x+\theta)^{\alpha-1}$$

のような形になる（x は区間 $[\theta, 1+\theta]$ の外では 0）。

分布族 S は θ を座標系として n 次元の多様体をなす。1点 θ における S の接空間を T_θ とする。T_θ の自然基底を $\{\partial_i\}$ とし，∂_i の確率変数による表現を

$$\partial_i \sim \partial_i l(x, \theta), \qquad \partial_i = \frac{\partial}{\partial \theta^i}$$

$$l(x, \theta) = \log p(x, \theta)$$

とする。T_θ のベクトル

$$\mathbb{A} = A^i \partial_i$$

は確率変数

$$A(x) = A^i \partial_i l(x, \theta)$$

で表現できる。

二つの確率分布 $p(x, \theta)$ と $p(x, \theta')$ の間の分離度を

Hellinger 距離

$$D(\theta, \theta') = 4 \int \left(\sqrt{p(x,\theta)} - \sqrt{p(x,\theta')}\right)^2 dx$$

または竹内の導入した

$$D(\theta, \theta') = -8 \log \int \sqrt{p(x,\theta)\, p(x,\theta')}\, dx$$

により定義しよう。このとき，$d\theta$ を微小とすると

$$D(\theta, \theta + d\theta) = \begin{cases} O(|d\theta|^2), & \alpha > 2 \\ |d\theta|^2 \log |d\theta|, & \alpha = 2 \\ |d\theta|^\alpha & 1 \leq \alpha < 2 \end{cases}$$

となることがわかっている。複雑な議論をさけるため，とりあえず $1 < \alpha < 2$ および $\alpha > 2$ の正則の場合の二つを考えよう。

T_θ における接ベクトル $\mathbb{h} = h^i \partial_i$ $(h(x) = h^i \partial_i l)$ のノルムを基本関数

$$F(\mathbb{h}, \theta) = \lim_{\varepsilon \to 0} \frac{1}{\varepsilon} \left\{ D(\theta, \theta + \varepsilon \mathbb{h}) \right\}^{\frac{1}{\alpha}}$$

$F(\mathbb{h}, \theta)$ は $\mathbb{h}$ について1次斉次の正関数である。これより，計量テンソル（フィンスラー計量）

$$g_{ij}(\mathbb{h}, \theta) = \frac{1}{2} \frac{\partial^2 F^2(\mathbb{h}, \theta)}{\partial h^i \partial h^j}$$

が定義できるから，S は Finsler 空間になる。

直積に分解できる場合，すなわち

$$p(x, \theta) = p_1(x_1, \theta_1)\, p_2(x_2, \theta_2),$$

$x = (x_1, x_2)$，　$\theta = (\theta_1, \theta_2)$　となる場合は，明らかに

$$F(\mathbb{h}, \theta) = \left[\{F_1(\mathbb{h}_1, \theta_1)\}^\alpha + \{F_2(\mathbb{h}_2, \theta_2)\}^\alpha \right]^{1/\alpha}$$

となる。ただし $\mathbb{h} = (\mathbb{h}_1, \mathbb{h}_2)$ で，F_1，F_2 はそれぞれ成分系の基本関数である。このとき二つのベクトル

$$(\mathbb{h}_1, 0)\ ,\quad (0, \mathbb{h}_2)$$

は互にどちらからみても直交していることがわかる。

同一独立分布の場合，すなわち

$$p(x_1, \cdots, x_m ; \theta) = \prod_{i=1}^{m} p(x_i, \theta)$$

の場合は，m 個の観測に基づく基本関数 F_m は

$$F_m(\mathbb{h}, \theta) = m^{\frac{1}{\alpha}} F(\mathbb{h}, \theta)$$

である。

次に，確率変数 $\mathbb{z} = (z_i)$，

$$z_i = \partial_i l(x, \theta)$$

に注目しよう。（点θは固定し，θを書くのを省略することがある。）明らかに

$$E[z_i] = 0$$

である。さらに，独立な観測 $x_1, \cdots, x_k$ に対して

$$k^{-\frac{1}{\alpha}} \sum_{i=1}^{k} z_i$$

は指数αの安定分布へ収束する。これをみるには，1次元の場合の（θ=0において）

$$z = l'(x) = \frac{c}{x}$$

の分布を調べて，z の分布 $g(z)$ は $z \to \infty$ のときに

$$g(z)dz = p(x)dx = z^{-\alpha-1}dz,$$

したがって

$$1 - P(z) = \int_z^\infty g(z)dz = z^{-\alpha},$$

これより（Feller, p.303）l' は指数αの安定分布の domain of attraction に入ることがわかる。同様の理由で行列 $\partial_i \partial_j l(x, \theta)$ は指数 $\alpha/2$ の安定分布の d. of attraction に入る。

z の収束する安定分布の特性関数を $\varphi(t, \theta)$ とすると

$$\varphi(t, \theta) = E[\exp\{i\, t^j \partial_j l(x, \theta)\}]$$

を用いて求まる。この安定分布（Z 自身は安定分布でないが）が定義する Minkowski 計量を $\hat{F}(t, \theta)$ としよう。$\hat{F}$ と F との関係を知る必要がある。

雑な議論を進めると、

$$\begin{aligned}\varphi(\varepsilon t) &= E[\exp\{i\varepsilon t^j \partial_j l(x, \theta)\}] \\ &= E[\exp\{i\Delta_{\varepsilon t} l(x, \theta)\}] \\ &= E[1 - \tfrac{1}{2}(\Delta_{\varepsilon t} l)^2] \\ &= \exp\{-\tfrac{1}{2}E[(\Delta_{\varepsilon t} l)^2]\},\end{aligned}$$

ここで

$$\Delta_{\varepsilon t} l(x, \theta) = l(x, \theta + \varepsilon t) - l(x, \theta),$$

一方

$$\begin{aligned}D(\theta, \theta + \varepsilon t) &= -8\log\int\left\{p + \frac{1}{2}\Delta_{\varepsilon t} p - \frac{1}{8}\frac{(\Delta p)^2}{p}\right\}dx \\ &= E[\{\Delta_{\varepsilon t} l(x, \theta)\}^2] = -E[\Delta^2_{\varepsilon t} l(x, \theta)] \\ &= \varepsilon^\alpha F(t)^{\frac{\alpha}{2}}\end{aligned}$$

が成立している。ただし $\Delta_{\varepsilon t} l$ についての期待値をとるときは $p(x, \theta)$ と $p(x, \theta + \varepsilon t)$ の共通の台上でのみ積分をとるものとする。これより F と $\hat{F}$ とが同じようなものであることがわかる。

ここで，m.l.e. $\hat{\theta}$ について少し考える。

$$\hat{\theta} = \theta_0 + d\theta$$

とおくと（θ_0 は真の値），N 個の測定 $x_1, \cdots, x_N$ に対し，

$$\sum_{k=1}^{N} \partial_i l(x_k, \theta_0 + d\theta) = 0$$

これより

$$N^{-1/\alpha} \sum \partial_i l(x_k, \theta_0) = -N^{-2/\alpha} \sum \partial_i \partial_j l(x_k, \theta_0) \widetilde{d\theta}^j$$

$$\widetilde{d\theta}^j = N^{1/\alpha} d\theta^j$$

左辺は確率変数 z_i に，右辺の係数項は w_{ij} に収束するから

$$\widetilde{d\theta}^j = w^{ji} z_i$$

となって $d\theta$ が $N^{-1/\alpha}$ のオーダーで収束することがわかる。確率変数 w_{ji} の性質を調べる必要がある。（正則な場合はこれは定数行列 g_{ij} に収束する。）

ここで $d\theta$ が双対ベクトルであり，かつ確率変数として扱われることに注意を要する。多分漸近的には

$$E[F(\widetilde{d\theta})] \geq n \quad \text{or} \quad \{E[F(\widetilde{d\theta})^{\alpha}]\}^{\frac{1}{\alpha}} \geq n$$

で，efficient な estimator に対しては等号が成立するといったものが Cramer-Rao の定理の非正則版として成立するのであろう。この際，等号を漸近的に達成するの

は必ずしも m.l.e. ではなくて、何らかの L_α, $L_{\alpha'}$ ノルムを用いて得られる推定量であるかもしれない。

文献

H. Rund, The Differential Geometry of Finsler Spaces, Springer-Verlag, Berlin-Göttingen-Heidelberg, 1959.

竹内啓, non-regular な統計的モデルの幾何学,「統計的推論の漸近理論と幾何学的構造」研究会資料 1983年10月